JOURNAL OF NUMBER THEORY 6, 448–480 (1974)

A Lower Bound for the Number of Solutions of Equations over Finite Fields*

WOLFGANG M. SCHMIDT

Department of Mathematics, University of Colorado, Boulder, Colorado 80302

Received October 1, 1973

DEDICATED TO PROFESSOR K. MAHLER ON THE OCCASION OF HIS 70TH BIRTHDAY

Let $f(X_1,\ldots, X_n)$ be an absolutely irreducible polynomial with coefficients in a finite field. Elementary methods are used to derive an explicit lower bound for the number of zeros of f.

1. INTRODUCTION

Our main result is the following

THEOREM. *Suppose $f(X_1,\ldots, X_m)$ is an absolutely irreducible polynomial of total degree $d > 0$, with coefficients in the finite field F_q with q elements. Let A be the number of solutions $(x_1,\ldots, x_n)$ with components in F_q of the equation*

$$f(x_1,\ldots, x_n) = 0. \tag{1}$$

Suppose

$$q > 10^4 n^3 d^5 P^3([4 \log d]), \tag{2}$$

where [ ] denotes integer parts and $P(1) = 2$, $P(2) = 3,\ldots$ is the sequence of primes.[1] *Then*

$$A > q^{n-1} - (d-1)(d-2)\, q^{n-(3/2)} - 6d^2 q^{n-2}. \tag{3}$$

A solution $(x_1,\ldots, x_n)$ of (1) is called *non-singular* if not all the partial derivatives of f vanish at $(x_1,\ldots, x_n)$. In §3 we shall deduce the

COROLLARY. *Suppose f is not a polynomial in $X_1^p,\ldots, X_n^p$ where p is the characteristic, and suppose f and q satisfy the hypotheses of the Theorem. Then there are non-singular solutions.*

* Written with partial support from NSF Grant NSF-GP-33026X.

[1] In particular, $P(x) \sim x \log x$, and the right-hand side of (2) is $O(n^3 d^{5+\epsilon})$ for every $\epsilon > 0$.

It may be useful to discuss the background for our Theorem. When $n = 1$, the absolute irreducibility implies $d = 1$, so that $A = 1$, and (3) holds. Now suppose $n = 2$, i.e., $f(X_1, X_2)$ is a polynomial in two variables of degree $d > 0$, which is absolutely irreducible. Let A^* be the number of "finite" solutions of (1) as above, plus the number of "solutions at infinity." Let A^{**} be the number of prime divisors of degree 1 of the function field of the curve $f(X_1, X_2) = 0$. A. Weil in 1948 [14] used methods of algebraic geometry to show that

$$| A^{**} - q - 1 | \leqslant 2gq^{1/2}, \tag{4}$$

where g is the genus of the curve. Now

$$| A^{**} - A^* | \leqslant \sum (r_P - 1) \leqslant \sum r_P(r_P - 1),$$

where the sum is over the singular points P of the curve, and where r_P is the multiplicity of P. It is shown in algebraic geometry that

$$2g + \sum r_P(r_P - 1) \leqslant (d - 1)(d - 2),$$

and hence (4) yields

$$| A^* - q - 1 | \leqslant (d - 1)(d - 2)q^{1/2}.$$

Since $A \leqslant A^* \leqslant A + d$, we obtain

$$| A - q | < (d - 1)(d - 2)q^{1/2} + d. \tag{5}$$

Hence the Theorem is true if $n = 2$.

In a recent paper [7], I used the elementary method of S. A. Stepanov [8–12] to prove a result which implies that

$$| A - q | < 2d^3q^{1/2}, \tag{6}$$

provided that $q > 9(d + 1)^4$. By using the zeta function of the curve $f(X_1, X_2) = 0$, one finds that (6) implies (4) and hence (5). On the other hand, Bombieri [1] used ideas of Stepanov together with the zeta function to give a very elegant direct proof of (4).

Now suppose $n \geqslant 3$ and $f = f(X_1, \ldots, X_n)$. Let again A be the number of solutions of (1) with components in F_q, and let A^* be the number A of these "finite solutions," plus the number of "solutions at infinity." It is easily shown (see Lemma 2 below) that

$$| A^* - A | \leqslant d(q^{n-2} + \cdots + q + 1).$$

Hence if we are concerned with rough bounds, it matters little whether we estimate A or A^*. Estimates of A^* were given by S. Lang and A. Weil [5]

and by L. B. Nisnevich [6], with similar methods and similar results. S. Lang and A. Weil showed (in fact they dealt with a more general situation) that

$$| A^* - q^{n-1} | < (d-1)(d-2)\, q^{n-(3/2)} + c_1 q^{n-2}, \tag{7}$$

where $c_1 = c_1(d, n)$.

Their method is roughly as follows. Assume for simplicity that $n = 3$. Let P be a plane with parameter representation

$$X_i = a_{i1}T_1 + a_{i2}T_2 + a_{i3} \qquad (i = 1, 2, 3). \tag{8}$$

Substituting these values for X_1, X_2, X_3 into $f(X_1, X_2, X_3)$, we get a polynomial in the two variables T_1, T_2. Now it turns out that for "most" planes P, the polynomial in T_1, T_2 so obtained is absolutely irreducible. Hence by the case of two variables, the number $A^*(P)$ of solutions (finite and infinite) on the plane P satisfies

$$| A^*(P) - q - 1 | \leqslant (d-1)(d-2)q^{1/2},$$

for "most" planes P. This fact, together with an estimate of the number of exceptional planes P, yields (7). The coefficients a_{ij} in (8) belonging to exceptional planes satisfy a polynomial equation $h(a_{11}, \ldots, a_{33}) = 0$, with a polynomial h depending on f. Unfortunately, the degree of h is very large as a function of d, and hence the constant c_1 in (7) becomes very large; in fact it increases faster than exponentially with d. But it would not be difficult to give explicit (rather large) bounds for $c_1 = c_1(d, n)$. Also, it is easy to give a completely elementary version of the proof.

In view of the large size of c_1, the estimate (7) becomes useless when q is small. Our Theorem gives a reasonably good estimate (in particular, $A > 0$), provided only that (2) holds.

Now suppose that the hypersurface $f(X_1, \ldots, X_n) = 0$ is "nonsingular" and that d or q is odd. Then it follows from work of B. Dwork [3, 4] together with (7) that

$$| A^* - (q^{n-1} + \cdots + q + 1)| \leqslant c_2 q^{n-(3/2)},$$

where

$$c_2 = d^{-1}((d-1)^{n+1} + (-1)^{n+1}(d-1)).$$

This does not imply (3). But the Weil conjectures on non-singular hypersurfaces[2] would imply that

$$| A^* - (q^{n-1} + \cdots + q + 1)| \leqslant c_2 q^{(n-1)/2},$$

[2] This conjecture was proved recently by P. Deligne, in a non-elementary way.

which would of course give much more than (3). On the other hand, our Theorem does not require non-singularity.

Our method is elementary and consists of a development of ideas in [7], which in turn were inspired by the work of Stepanov. We will make use of (5). If instead of (5) one uses the strictly elementary estimate (6), then one obtains our Theorem with (2) replaced by both (2) and

$$q > 2000\, n^2 d^6, \tag{2'}$$

and (3) replaced by

$$A > q^{n-1} - 3d^3 q^{n-(3/2)}. \tag{3'}$$

Our method does not at present give an equally good upper bound for A, except in the case when the algebraic function $\mathfrak{Y}$ with $f(X_1, \ldots, X_{n-1}, \mathfrak{Y}) = 0$ is either normal or of degree $\leqslant 3$ over the field of rational functions in $X_1, \ldots, X_{n-1}$. We shall not deal with this upper bound. One can use our Theorem, together with birational transformations, to give lower bounds for the number of points on varieties defined over F_q. Again, we shall not deal with this in the present paper.

Many arguments used in the sequel appear in a simpler form in the preceding paper [7], and hence this paper may serve as an introduction to the present one.

2. Notation

We shall employ almost the same notation as in [7]. Throughout, p will be a prime and q a power of p. We shall write $\overline{F}_q$ for the algebraic closure of F_q, and, more generally, $\overline{K}$ for the algebraic closure of a field K. Elements of $\overline{F}_q$ will be denoted by $x, y, \ldots$. We shall write $X, Y, \ldots, X_1, X_2, \ldots$ for variables, and $\mathfrak{X}, \mathfrak{Y}, \ldots$ for algebraic functions, i.e., for quantities which are algebraically dependent on some of the variables $X, Y, \ldots$. Polynomials will be written as $a(X)$, $b(X)$, $a(X_1, \ldots, X_n), \ldots$. Unless stated otherwise, the degree of $a(X_1, \ldots, X_n)$, denoted $\deg a$, will mean the total degree. The notation

$$\deg_{X_i} a(X_1, \ldots, X_n)$$

will mean the degree of $a(X_1, \ldots, X_n)$ in the variable $X_i (i = 1, \ldots, n)$.

If K' is a finite algebraic extension of a field K, then $[K' : K]$ will denote the degree of that extension. $K(X_1, \ldots, X_u, \mathfrak{X}_1, \ldots, \mathfrak{X}_v, x_1, \ldots, x_w)$ will be the field obtained by adjoining $X_1, \ldots, x_w$ to the field K. Thus $K(X_1, \ldots, X_u)$ is the field of rational functions in u variables over K.

The number of elements of a finite set ω will be denoted by $|\omega|$.

3. Elementary Upper Bounds

Lemma 1. *Suppose $g(X_1,..., X_n)$ is a non-zero polynomial of degree d with coefficients in F_q. Then the number of solutions $(x_1,..., x_n)$ with $x_i \in F_q$ of*

$$g(x_1,..., x_n) = 0 \tag{9}$$

is at most dq^{n-1}. If g is homogeneous, then the number of solutions $(x_1,..., x_n) \neq (0,..., 0)$ is at most $d(q^{n-1} - 1)$.

Proof.[3] The lemma is obviously true if $d = 0$ or $d = 1$. It is also true if $n = 1$. Suppose $n > 1$, $d > 1$, and suppose the lemma is already proved for polynomials in $\leqslant n$ variables of degree $< d$, and for polynomials in $< n$ variables of degree $\leqslant d$.

Suppose at first that $g(X_1,..., X_n)$ is not divisible by $X_1 - x$ for any $x \in F_q$. Then for every x, $g(x, X_2,..., X_n)$ is a non-zero polynomial in $n - 1$ variables, hence by our assumption has $\leqslant dq^{n-2}$ solutions in $x_2,..., x_n$. Summing over $x \in F_q$, we obtain $\leqslant qdq^{n-2} = dq^{n-1}$ solutions. The number of solutions with $x_1 \neq 0$ is $\leqslant (q - 1)\, dq^{n-2}$. If g is homogeneous, then so is $g(0, X_2,..., X_n)$. The number of non-zero solutions of $g(0, x_2,..., x_n) = 0$ is $\leqslant d(q^{n-2} - 1)$. Thus for g homogeneous, the number of non-zero solutions is altogether

$$\leqslant (q - 1)\, dq^{n-2} + d(q^{n-2} - 1) = d(q^{n-1} - 1).$$

Now suppose that $g(X_1,..., X_n)$ is divisible by $X_1 - x$, say that $g(X_1,..., X_n) = (X_1 - x)\, g_1(X_1,..., X_n)$. The number of zeros is at most q^{n-1} plus the number of zeros of g_1, hence is $\leqslant q^{n-1} + (d - 1)q^{n-1} = dq^{n-1}$. If g is homogeneous, then it can only be divisible by $X_1 - 0 = X_1$, so that $g(X_1,..., X_n) = X_1 g_1(X_1,..., X_n)$. The number of non-zero solutions now is $\leqslant (q^{n-1} - 1) + (d - 1)(q^{n-1} - 1)$.

Now let

$$f(X_1,..., X_n) = \sum_{i_1+\cdots+i_n \leqslant d} a_{i_1 \cdots i_n} X_1^{i_1} \cdots X_n^{i_n}$$

be a polynomial of degree d with coefficients in F_q. As in the introduction, A will denote the number of solutions $(x_1,..., x_n)$ with $x_i \in F_q$ of (1). With f we associate the form

$$f^*(X_0, X_1,..., X_n) = \sum_{i_0+i_1+\cdots+i_n=d} a_{i_1 \cdots i_n} X_0^{i_0} X_1^{i_1} \cdots X_n^{i_n}.$$

A^* is defined as the number of solutions of $f^*(x_0, x_1,..., x_n) = 0$ in

[3] See also the lemma in [2], §5.2.

non-zero $(n+1)$-tuples $(x_0, x_1, \ldots, x_n)$ with components of F_q, where two $(n+1)$-tuples are considered equal if they are proportional.

LEMMA 2. *Suppose $n \geqslant 2$. Then*

$$| A^* - A | \leqslant d(q^{n-2} + \cdots + q + 1).$$

Proof. We have $A^* = A_0^* + A_1^*$, where A_0^* counts the solutions $(x_0, x_1, \ldots, x_n)$ with $x_0 \neq 0$, and A_1^* counts the solutions with $x_0 = 0$. Every $(n+1)$-tuple with $x_0 \neq 0$ is proportional to a unique $(n+1)$-tuple with $x_0 = 1$. Hence A_0^* is the number of $(x_1, \ldots, x_n)$ with $f^*(1, x_1, \ldots, x_n) = 0$, i.e., with $f(x_1, \ldots, x_n) = 0$, and we have $A_0^* = A$. Thus $A^* - A = A_1^*$, which is $(q-1)^{-1}$ times the number of $(x_1, \ldots, x_n) \neq (0, \ldots, 0)$ with $f^*(0, x_1, \ldots, x_n) \neq 0$, and by Lemma 1 we have

$$A^* - A = A_1^* \leqslant (q-1)^{-1} d(q^{n-1} - 1) = d(q^{n-2} + \cdots + q + 1).$$

LEMMA 3. *Suppose $u_1(X, Y)$, $u_2(X, Y)$ are polynomials of degrees e_1, e_2 with coefficients in some field K. Suppose u_1, u_2 have no common factor of positive degree. Then the number of solutions x, y in K of*

$$u_1(x, y) = u_2(x, y) = 0$$

is $\leqslant e_1 e_2$.

Proof. We may assume without loss of generality that K is infinite. Suppose there are at least λ common zeros. One can find a set of λ distinct parallel lines in the plane K^2, with each line containing precisely one of the solutions (x, y). Hence after a linear transformation of the variables, one obtains polynomials v_1, v_2 with degrees e_1, e_2, and with no common factor, which have common zeros $(x_1, y_1), \ldots, (x_\lambda, y_\lambda)$ with $x_1, \ldots, x_\lambda$ all distinct. We have $r(x_1) = \cdots = r(x_\lambda) = 0$, where $r(X)$ is the resultant of v_1, v_2 when considered as polynomials in Y. Since v_1, v_2 have no common factor, the resultant is a non-zero polynomial of degree at most $e_1 e_2$. Hence $\lambda \leqslant e_1 e_2$.

COROLLARY. *Suppose $u_0(X, Y)$, $u_1(X, Y), \ldots, u_t(X, Y)$ are polynomials of degree $\leqslant e$ which are relatively prime, i.e., which have no common factor of positive degree. Let all these polynomials have coefficients in some field K. Then the number of common zeros, i.e., the number of pairs x, y in K with*

$$u_0(x, y) = \cdots = u_t(x, y) = 0 \tag{10}$$

is $\leqslant e^2$.

Proof. Let $v(X, Y)$ be the greatest common divisor of $u_1, \ldots, u_t$, and

let $u_i(X, Y) = v(X, Y)\, w_i(X, Y)$ $(i = 1,..., t)$. If d is the degree of v, then deg $w_i \leqslant e - d$ $(i = 1,..., t)$. Every solution of (10) is either a solution of

$$u_0(x, y) = v(x, y) = 0 \tag{11}$$

or of

$$w_1(x, y) = \cdots = w_t(x, y) = 0. \tag{12}$$

Since u_0, v are relatively prime, the number of solutions of (11) is $\leqslant de$ by Lemma 3. Since $w_1,..., w_t$ are relatively prime, the number of solutions of (12) is $\leqslant (e - d)^2$, if we assume inductively that the Corollary is true for $t - 1$ in place of t. Since

$$de + (e - d)^2 = e^2 - de + d^2 \leqslant e^2,$$

the Corollary follows.

LEMMA 4. *Suppose* $u_0(X_1,..., X_n)$, $u_1(X_1,..., X_n),..., u_t(X_1,..., X_n)$ *are polynomials of degree* $\leqslant e$ *with coefficients in* F_q *and without a common factor. Then the number of solutions of*

$$u_0(x_1,..., x_n) = \cdots = u_t(x_1,..., x_n) = 0 \tag{13}$$

with components in F_q *is* $\leqslant 2ne^3q^{n-2}$.

Proof. We proceed by induction on n. When $n = 1$, there is no solution, and when $n = 2$, the assertion follows from the Corollary to Lemma 3. When $n \geqslant 3$, write A_1 for the number of solutions $(x_1,..., x_n)$ where all the polynomials

$$u_0(x_1,..., x_{n-1}, X_n),..., u_t(x_1,..., x_{n-1}, X_n) \tag{14}$$

are identically zero, and write A_2 for the number of remaining solutions. Then A_1 equals q times the number of common zeros of the polynomials $v_{ij}(X_1,..., X_{n-1})$ $(0 \leqslant i \leqslant t, 0 \leqslant j \leqslant e)$, where v_{ij} is the coefficient of X_n^j in $u_i(X_1,..., X_n)$. Since the polynomials v_{ij} have no common factor of positive degree, the number of their common zeros $(x_1,..., x_{n-1})$ is $\leqslant 2(n - 1)e^3q^{n-3}$ by induction, and $A_1 \leqslant 2(n - 1)e^3q^{n-2}$.

Now let $w_1,..., w_s$ be the distinct irreducible factors of u_0. For each w_i there is a j_i such that $1 \leqslant j_i \leqslant t$ and u_{j_i} is not divisible by w_i. Put

$$v_i = u_{j_i} w_1 \cdots w_{i-1} w_{i+1} \cdots w_s \qquad (i = 1,..., s),$$

$$v = v_1 + \cdots + v_s.$$

Then v is of degree $\leqslant 2e$, and u_0, v have no common factor. Every

solution of (13) is a common zero of u_0 and v. Hence it satisfies $r(x_1, \ldots, x_{n-1}) = 0$, where r is the resultant of u_0, v when considered as polynomials in X_n. This resultant is a non-zero polynomial of degree $\leqslant 2e^2$, so that the number of possibilities for $x_1, \ldots, x_{n-1}$ is $\leqslant 2e^2q^{n-2}$ by Lemma 1. Now for A_2 we are concerned with solutions where the polynomials (14) are not all zero. Hence for given $x_1, \ldots, x_{n-1}$, there are at most e possibilities for x_n. Hence $A_2 \leqslant 2e^3q^{n-2}$. Thus $A_1 + A_2 \leqslant 2ne^3q^{n-2}$, and the lemma is true.

Proof of the Corollary to the Theorem. Because of our special assumption on f, not all the partial derivatives are identically zero. Let $g(X_1, \ldots, X_n)$ be one of the non-zero derivatives. Every "singular" solution of (1) is a common zero of f and g, and by Lemma 4 there are $\leqslant 2nd^3q^{n-2}$ such zeros. The number of non-singular solutions is therefore $\geqslant A - 2nd^3q^{n-2}$, and in particular it is positive by (2) and (3).

4. The Number of Points on a (not necessarily Irreducible) Surface

If $u(X, Y)$ is an absolutely irreducible polynomial of degree $d > 0$ with coefficients in F_q, then the number A of zeros of u in F_q satisfies

$$| A - q | < \alpha(q, d), \tag{15}$$

where we may take

$$\alpha = \alpha_1 = (d - 1)(d - 2)\, q^{1/2} + d$$

by (5). Or, a reader who prefers the strictly elementary estimate (6), may take

$$\alpha = \alpha_2 = 2d^3q^{1/2},$$

provided that $q > 9(d + 1)^4$. In either case, the function $\alpha(q, d)$ has

$$\alpha(q, d) + \alpha(q, d') \leqslant \alpha(q, d + d'). \tag{16}$$

Lemma 5. *Suppose $f(X, Y)$ is a polynomial of degree d with coefficients in F_q, and not necessarily irreducible. There is an integer v with*

$$0 \leqslant v \leqslant d,$$

such that the number A of zeros of f over F_q satisfies

$$| A - vq | < \alpha(q, d) + d^2.$$

Proof. We may write

$$f = cf_1^{\varphi_1} \cdots f_k^{\varphi_k},$$

where c is a constant and no two of the polynomials $f_1, \ldots, f_k$ are proportional to each other, and where each polynomial f_i has coefficients in $\bar{F}_q$ and is irreducible over $\bar{F}_q$, and has some coefficient equal to 1. Suppose precisely ν of $f_1, \ldots, f_k$ have all their coefficients in F_q, say $f_1, \ldots, f_\nu$. Denote the degree of f_i by d_i, and let A_i be the number of zeros of f_i with components in F_q $(i = 1, \ldots, k)$. Let A_{ij} be the number of common zeros of f_i and f_j. We have

$$| A_i - q | < \alpha(q, d_i) \qquad (i \leqslant 1 \leqslant \nu)$$

by (15) and

$$A_{ij} \leqslant d_i d_j \qquad (1 \leqslant i < j \leqslant \nu)$$

by Lemma 3.

Now for $i > \nu$ we have

$$f_i = f_{i0} + \alpha_{i1} f_{i1} + \cdots + \alpha_{it_i} f_{it_i},$$

where the f_{ih} are polynomials with coefficients in F_q, where $\alpha_{i1}, \ldots, \alpha_{it_i}$ are in $\bar{F}_q$ with $1, \alpha_{i1}, \ldots, \alpha_{it_i}$ linearly independent over F_q, and where $t_i \geqslant 1$. Moreover, the polynomials $f_{i0}, \ldots, f_{it_i}$ have no common factor. Every zero of f_i with components in F_q is a common zero of $f_{i0}, \ldots, f_{it_i}$. By the Corollary to Lemma 3 there are at most d_i^2 such common zeros, i.e.,

$$A_i \leqslant d_i^2 \qquad (\nu < i \leqslant k).$$

The number A of zeros of the given polynomial f has

$$\begin{aligned} A &\geqslant \sum_{i=1}^{\nu} A_i - \sum_{\substack{i,j=1 \\ i \neq j}}^{\nu} A_{ij} \\ &\geqslant \nu q - \sum_{i=1}^{\nu} \alpha(q, d_i) - \sum_{i \neq j} d_i d_j \\ &> \nu q - \alpha(q, d) - d^2 \end{aligned}$$

by (16). On the other hand,

$$\begin{aligned} A &\leqslant \sum_{i=1}^{\nu} A_i + \sum_{i=\nu+1}^{k} A_i \leqslant \nu q + \sum_{i=1}^{\nu} \alpha(q, d_i) + d_{\nu+1}^2 + \cdots + d_k{}^2 \\ &< \nu q + \alpha(q, d) + d^2. \end{aligned}$$

Now suppose $n \geqslant 3$. We shall denote 2-dimensional planes (not necessarily through the origin) of n-dimensional space F_q^n by P. If ω_0 is the number of planes, ω_1 the number of planes through a given point and ω_2 the number of planes through two given distinct points, then

$$\omega_0 = q^n(q^n - 1)(q^n - q)\, q^{-2}(q^2 - 1)^{-1}\,(q^2 - q)^{-1},$$
$$\omega_1 = (q^n - 1)(q^n - q)(q^2 - 1)^{-1}\,(q^2 - q)^{-1},$$
$$\omega_2 = (q^n - q)(q^2 - q)^{-1}.$$

Let $A(P)$ be the number of zeros of a polynomial $f(X_1, \ldots, X_n)$ which lie on a plane P, and, as usual, let A be the total number of zeros. We shall denote zeros by (x), (y),... .

LEMMA 6. *We have*

$$\sum_P (A(P) - Aq^{2-n})^2 \leqslant d\omega_1 q^{n-1},$$

where d is the degree of f.

Proof.

$$\sum_P A(P) = \sum_{(x)} \sum_{P \ni (x)} 1 = \sum_{(x)} \omega_1 = A\omega_1,$$

so that the mean value of $A(P)$ over all planes P is $A\omega_1/\omega_0 = Aq^{2-n}$. Next,

$$\sum_P A(P)^2 = \sum_{(x)} \sum_{(y)} \sum_{P \ni (x),(y)} 1 = \sum_{(x) \neq (y)} \omega_2 + \sum_{(x)} \omega_1$$
$$= A(A - 1)\,\omega_2 + A\omega_1 \leqslant A^2\omega_2 + A\omega_1.$$

Thus

$$\sum_P (A(P) - Aq^{2-n})^2 = \Big(\sum_P A^2(P)\Big) - \omega_0 A^2 (q^{2-n})^2$$
$$\leqslant A^2\omega_2 + A\omega_1 - A^2\omega_1^2/\omega_0 \leqslant A\omega_1 \leqslant d\omega_1 q^{n-1},$$

since $\omega_0\omega_2 < \omega_1^2$, and since $A \leqslant dq^{n-1}$ by Lemma 1.

LEMMA 7. *Suppose*

$$q > 6d^2 q^{1/2} + 4\alpha(q, d). \tag{17}$$

Then there is an integer v_0, $0 \leqslant v_0 \leqslant d$, with

$$|A - v_0 q^{n-1}| < (\alpha(q, d) + 5d^2)\, q^{n-2}.$$

Proof. We have $\sum_P (dq) = \omega_0 dq = \omega_1 dq^{n-2}$, and hence by Lemma 6 there is a plane P with $| A(P) - Aq^{2-n} | \leqslant \sqrt{dq}$. By Lemma 5 there is an integer ν_0 with

$$| A(P) - \nu_0 q | < \alpha(q, d) + d^2.$$

Here $0 \leqslant \nu_0 \leqslant d$ if f does not vanish identically on P, but $A(P) = q^2$ and $\nu_0 = q$ if f is identically zero on P. Combining our inequalities we obtain

$$| A - \nu_0 q^{n-1} | < (\alpha(q, d) + d^2 + (dq)^{1/2})\, q^{n-2}.$$

If we had $\nu_0 = q$, then $A > q^n - \alpha(q, d) q^{n-2} - 2d^2 q^{n-(3/2)}$, and on the other hand $A \leqslant dq^{n-1}$ by Lemma 1. These two inequalities together with (17) are incompatible. Thus $0 \leqslant \nu_0 \leqslant d$.

But this is not quite the end of the story: Let Π_j be the set of planes P with $| \nu(P) - \nu_0 | = j$. Planes $P \in \Pi_j$ with $j > 0$ have $A(P) - \nu_0 q = (\nu(P) - \nu_0)\, q + A(P) - \nu(P) q$, whence

$$| A(P) - \nu_0 q | < jq + \alpha(q, d) + d^2 < (j+1) q \leqslant 2jq$$

by (17). On the other hand, again by (17),

$$\begin{aligned} | A(P) - Aq^{2-n} | &\geqslant | A(P) - \nu_0 q | - q^{2-n} | A - \nu_0 q^{n-1} | \\ &> jq - \alpha(q, d) - d^2 - \alpha(q, d) - 2d^2 q^{1/2} > \tfrac{1}{2} jq. \end{aligned}$$

Thus Lemma 6 yields

$$\tfrac{1}{4} q^2 (| \Pi_1 | + 2^2 | \Pi_2 | + \cdots + q^2 | \Pi_q |) \leqslant d\omega_1 q^{n-1},$$

whence

$$| \Pi_1 | + 2 | \Pi_2 | + \cdots + q | \Pi_q | \leqslant 4 d \omega_1 q^{n-3}.$$

We obtain

$$\begin{aligned} | A - \nu_0 q^{n-1} | &= \omega_1^{-1} \Big| \sum_P (A(P) - \nu_0 q) \Big| \leqslant \omega_1^{-1} \sum_P | A(P) - \nu_0 q | \\ &\leqslant \omega_1^{-1} \Big(\omega_0 (\alpha(q, d) + d^2) + \sum_{j=1}^{q} 2jq\, | \Pi_j | \Big) \\ &\leqslant q^{n-2} (\alpha(q, d) + d^2 + 8d) \\ &\leqslant q^{n-2} (\alpha(q, d) + 5d^2), \end{aligned}$$

if $d \geqslant 2$. The lemma is obviously true if $d = 1$.

5. Certain Field Extensions

LEMMA 8. *Suppose $f(X, Y)$, $g(X, Y)$ are absolutely irreducible polynomials with coefficients in some field K, of respective degrees $d > 0$, $e > 0$ in Y. Let X, Z be variables, and $\mathfrak{Y}$, $\mathfrak{U}$ algebraic functions with*

$$f(X, \mathfrak{Y}) = 0, \qquad g(Z, \mathfrak{U}) = 0.$$

Then

$$[K(X, Z, \mathfrak{Y}, \mathfrak{U}) : K(X, Z)] = de.$$

Proof. The case $f = g$ was proved as Hilfssatz 2 in [7]. The general case is proved in the same way.[4]

Now suppose that K is of characteristic p, and that q is a power of p. Given a polynomial $f(X, Y) = \sum a_{ij}X^iY^j$ with coefficients in K, write $f^{[q]}(X, Y) = \sum a_{ij}^q X^iY^j$. Now the mapping $x \to x^q$ is an automorphism of $\overline{K}$, and hence if f is absolutely irreducible, then so is $f^{[q]}$. We therefore have the

COROLLARY. *Suppose $f(X, Y)$ is an absolutely irreducible polynomial of degree $d > 0$ in Y with coefficients in a field K of characteristic p. Let X, Z be variables and $\mathfrak{Y}$, $\mathfrak{U}$ algebraic functions with*

$$f(X, \mathfrak{Y}) = 0, \qquad f^{[q]}(Z, \mathfrak{U}) = 0.$$

Then

$$[K(X, Z, \mathfrak{Y}, \mathfrak{U}) : K(X, Z)] = d^2.$$

LEMMA 9. *Let $r(Y)$, $s(Y)$, $g_r(X, Y)$, $g_s(X, Y)$ be polynomials with coefficients in a field K, with g_r, g_s absolutely irreducible and of respective degrees $d_r > 0$, $d_s > 0$ in Y, and suppose the degree of*

$$r(Y) - s(Y)$$

is positive and prime to d_rd_s. Put

$$X^{(r)} = X + r(Y), \qquad X^{(s)} = X + s(Y), \tag{18}$$

and let $\mathfrak{Z}_r$, $\mathfrak{Z}_s$ satisfy

$$g_r(X^{(r)}, \mathfrak{Z}_r) = 0, \qquad g_s(X^{(s)}, \mathfrak{Z}_s) = 0.$$

Then

$$[K(X, Y, \mathfrak{Z}_r, \mathfrak{Z}_s) : K(X, Y)] = d_rd_s.$$

[4] Rather than to look up [7], the reader might prefer to give his own proof of this simple lemma.

Proof. Since $X^{(r)}$, $X^{(s)}$ are algebraically independent, Lemma 8 implies that $[K(X^{(r)}, X^{(s)}, \mathfrak{Z}_r, \mathfrak{Z}_s) : K(X^{(r)}, X^{(s)}] = d_r d_s$. The polynomial

$$u(Z) = r(Z) - s(Z) - X^{(r)} + X^{(s)}$$

is irreducible over $K(X^{(r)}, X^{(s)})$ and of a degree e prime to $d_r d_s$. Now $u(Y) = 0$ by (18), so that Y is algebraic of degree e over $K(X^{(r)}, X^{(s)})$. Since e is prime to $d_r d_s$, we get

$$[K(X^{(r)}, X^{(s)}, Y, \mathfrak{Z}_r, \mathfrak{Z}_s) : K(X^{(r)}, X^{(s)}, Y)] = d_r d_s .$$

Since $K(X^{(r)}, X^{(s)}, Y) = K(X, Y)$, the lemma follows.

6. A Linear Independence Result

We shall employ the notation $s_1(U_1, \ldots, U_d) = -(U_1 + \cdots + U_d), \ldots, s_d(U_1, \ldots, U_d) = (-1)^d U_1 \cdots U_d$ for the elementary symmetric polynomials in d variables $U_1, \ldots, U_d$.

Lemma 10. *Suppose $a(U_1, \ldots, U_d)$ is a symmetric polynomial in $U_1, \ldots, U_d$. Then there exists a polynomial $b(V_1, \ldots, V_d)$ such that*

$$a(U_1, \ldots, U_d) = b(s_1(U_1, \ldots, U_d), \ldots, s_d(U_1, \ldots, U_d)).$$

If $a(U_1, \ldots, U_d)$ has degree t in each variable U_i, then $b(V_1, \ldots, V_d)$ has total degree t.

Proof. The first assertion is well-known. The assertion about the degree follows, e.g., from the proof given in Van der Waerden [13].

Lemma 11. *Suppose*

$$f(X, Y) = Y^d + g_1(X)Y^{d-1} + \cdots + g_d(X),$$

where $g_i(X)$ is a polynomial with coefficients in a field K of characteristic p and with

$$\deg g_i(X) \leqslant i\psi \qquad (i = 1, \ldots, d). \tag{19}$$

Suppose f is absolutely irreducible, and let $\mathfrak{Y}$ satisfy $f(X, \mathfrak{Y}) = 0$. Then the quantities

$$\mathfrak{Y}^i \mathfrak{Y}^{qj} X^k \qquad (0 \leqslant i, j \leqslant d - 1; 0 \leqslant k < (q/d) - d\psi)$$

are linearly independent over the field $K(X^q)$.

Proof. Although the lemma is very similar to Hilfssatz 3 of [7], it is probably most convenient to give a complete proof here. Let $a(X, Y, Z, W)$ be a polynomial with coefficients in K with

$$\deg_X a < (q/d) - d\psi, \qquad \deg_Y a \leqslant d - 1, \qquad \deg_W a \leqslant d - 1.$$

We have to verify that

$$a(X, \mathfrak{Y}, X^q, \mathfrak{Y}^q) = 0 \qquad \textit{implies that} \quad a(X, Y, Z, W) = 0.$$

We have $\mathfrak{Y}^d = -g_1(X)\mathfrak{Y}^{d-1} - \cdots - g_d(X)$, and using (19) and induction on t, we obtain

$$\mathfrak{Y}^{d-1+t} = g_1^{(t)}(X)\,\mathfrak{Y}^{d-1} + \cdots + g_d^{(t)}(X) \qquad (t = 1, 2, \ldots), \tag{20}$$

with

$$\deg g_i^{(t)}(X) \leqslant (t + i - 1)\psi \leqslant (t + d - 1)\psi \qquad (i = 1, \ldots, d). \tag{21}$$

Put

$$a(X, Y, Z; W_1, \ldots, W_d) = \prod_{h=1}^{d} a(X, Y, Z, W_h).$$

The polynomial $a(X, Y, Z; W_1, \ldots, W_d)$ is symmetric in $W_1, \ldots, W_d$ and of degree $\leqslant d - 1$ in each W_i, so that by Lemma 10 we obtain

$$a(X, Y, Z; W_1, \ldots, W_d) = b(X, Y, Z; s_1(W_1, \ldots, W_d), \ldots, s_d(W_1, \ldots, W_d)), \tag{22}$$

where $b(X, Y, Z; V_1, \ldots, V_d)$ is a polynomial of total degree $\leqslant d - 1$ in the variables $V_1, \ldots, V_d$. Now since b has degree $\leqslant d(d-1) = d - 1 + (d-1)^2$ in Y, an application of (20) with $t = (d-1)^2$ shows that

$$b(X, \mathfrak{Y}, Z; V_1, \ldots, V_d) = c(X, \mathfrak{Y}, Z; V_1, \ldots, V_d), \tag{23}$$

where $c(X, Y, Z; V_1, \ldots, V_d)$ is a polynomial with

$$\deg_Y c \leqslant d - 1, \tag{24}$$

$$\deg_X c < d((q/d) - d\psi) + ((d-1)^2 + d - 1)\psi < q. \tag{25}$$

Now let $\mathfrak{Y}_1 = \mathfrak{Y}, \mathfrak{Y}_2, \ldots, \mathfrak{Y}_d$ be the roots of the polynomial $f(X, Y)$ in Y, i.e., suppose that

$$f(X, Y) = (Y - \mathfrak{Y}_1) \cdots (Y - \mathfrak{Y}_d).$$

Then $g_j(X) = s_j(\mathfrak{Y}_1,\ldots, \mathfrak{Y}_d)$, and raising this to the q-th power, we obtain

$$g_j^{[q]}(X^q) = s_j(\mathfrak{Y}_1{}^q,\ldots, \mathfrak{Y}_d{}^q) \qquad (j = 1,\ldots, d). \tag{26}$$

Now suppose $a(X, \mathfrak{Y}, X^q, \mathfrak{Y}^q) = 0$. Then also $a(X, \mathfrak{Y}, X^q; \mathfrak{Y}_1{}^q,\ldots, \mathfrak{Y}_d{}^q) = 0$, whence in view of (22) and (26),

$$b(X, \mathfrak{Y}, X^q; g_1^{[q]}(X^q),\ldots, g_d^{[q]}(X^q)) = 0,$$

whence by (23),

$$c(X, \mathfrak{Y}, X^q; g_1^{[q]}(X^q),\ldots, g_d^{[q]}(X^q)) = 0.$$

Since $\mathfrak{Y}$ is of the degree d over $K(X)$, (24) yields

$$c(X, Y, X^q; g_1^{[q]}(X^q),\ldots, g_d^{[q]}(X^q)) = 0.$$

This is an identity in X, Y. We now substitute $X_1 + X_2$ for X and obtain

$$c(X_1 + X_2, Y, X_1{}^q + X_2{}^q; g_1^{[q]}(X_1{}^q + X_2{}^q),\ldots, g_d^{[q]}(X_1{}^q + X_2{}^q)) = 0. \tag{27}$$

The left hand side of this equation equals

$$c(X_1 + X_2, Y, X_1{}^q; g_1^{[q]}(X_1{}^q),\ldots, g_d^{[q]}(X_1{}^q)) + X_2{}^q d(X_1, X_2, Y),$$

where $d(X_1, X_2, Y)$ is a polynomial. Now by virtue of (25), $c(X_1+X_2, Y, X_1{}^q; g_1^{[q]}(X_1{}^q),\ldots)$ is of degree $< q$ in X_2, and it follows that $c(X_1 + X_2, Y, X_1{}^q; g_1^{[q]}(X_1{}^q),\ldots, g_d^{[q]}(X_1{}^q)) = 0$. Since $X_1 + X_2$, Y, $X_1{}^q$ are algebraically independent, we obtain the identity

$$c(X, Y, Z; g_1^{[q]}(Z),\ldots, g_d^{[q]}(Z)) = 0$$

in the three variables X, Y, Z.

We now substitute $\mathfrak{Y}$ for Y and we obtain

$$b(X, \mathfrak{Y}, Z; g_1^{[q]}(Z),\ldots, g_d^{[q]}(Z)) = 0, \tag{28}$$

by (23). Now let $\mathfrak{U}_1,\ldots, \mathfrak{U}_d$ be the roots of $f^{[q]}(Z, U)$ in U, i.e., suppose that

$$f^{[q]}(Z, U) = (U - \mathfrak{U}_1)\cdots(U - \mathfrak{U}_d).$$

Then $g_i^{[q]}(Z) = s_i(\mathfrak{U}_1,\ldots, \mathfrak{U}_d)$ $(i = 1,\ldots, d)$. Substituting these values into (28) and recalling the definition of $b(X, Y, Z; \ldots)$, we get

$a(X, \mathfrak{Y}, Z; \mathfrak{U}_1, ..., \mathfrak{U}_d) = 0$. Thus there is a j, $1 \leqslant j \leqslant d$, such that $\mathfrak{U} = \mathfrak{U}_j$ has

$$a(X, \mathfrak{Y}, Z, \mathfrak{U}) = 0.$$

The quantities $\mathfrak{Y}$, $\mathfrak{U}$ are as in the Corollary to Lemma 8. Since $a(X, Y, Z, W)$ has degree $\leqslant d - 1$ in Y and in W, we obtain

$$a(X, Y, Z, W) = 0.$$

7. A Special Hypothesis

In what follows, we shall assume that $n \geqslant 3$, and put

$$m = n - 1. \tag{29}$$

We shall write the polynomial of the Theorem as $f(X_1, ..., X_m, Y)$. Since the mapping $x \to x^p$ is an automorphism of F_q, the number of zeros of a polynomial $g(X_1, ..., X_m, Y)$ is the same as the number of zeros of $g(X_1, ..., X_m, Y^p)$. Hence we may suppose without loss of generality that $f(X_1, ..., X_m, Y)$ is not a polynomial in Y^p, i.e., that it is separable in Y.

Lemma 12. *Suppose $f(X_1, ..., X_m, Y)$ is a polynomial of degree $d > 0$ with coefficients in F_q and separable in Y. Suppose $q > 2d$. There exist c, $a_1, ..., a_m$ in F_q such that the polynomial*

$$g(X_1, ..., X_m, Y) = cf(X_1 + a_1 Y, ..., X_m + a_m Y, Y)$$

is again separable in Y, and has the term Y^d with coefficient 1.

Proof. We have

$$g(0, ..., 0, Y) = cf(a_1 Y, ..., a_m Y, Y) = ch_1(a_1, ..., a_m) Y^d + \cdots,$$

where h_1 is a polynomial of degree $\leqslant d$. Since f is of (precise) degree d, the polynomial h_1 is non-zero.

Since f is separable in Y, it contains a term $X_1^{i_1} \cdots X_m^{i_m} Y^j$ with $p \nmid j$ with a non-zero coefficient. The term $X_1^{i_1} \cdots X_m^{i_m} Y^j$ occurs in g with a coefficient $ch_2(a_1, ..., a_m)$, where h_2 is a non-zero polynomial of degree $\leqslant d$.

We have to choose $a_1, ..., a_m$ so that $h_1(a_1, ..., a_m)\, h_2(a_1, ..., a_m) \neq 0$. By Lemma 1, the number of $a_1, ..., a_m$ in F_q with $h_1 h_2(a_1, ..., a_m) = 0$ is

$\leqslant 2dq^{m-1} < q^m$. Hence there are $a_1, \ldots, a_m$ with the desired property. We finally pick c such that $ch_1(a_1, \ldots, a_m) = 1$.

In view of the lemma just proved, we may assume that the polynomial of the Theorem is of the type

$$f(X_1, \ldots, X_m, Y) = Y^d + g_1(X_1, \ldots, X_m)Y^{d-1} + \cdots + g_d(X_1, \ldots, X_m), \quad (30)$$

where g_i is a polynomial with

$$\deg g_i \leqslant i \qquad (i = 1, \ldots, d). \quad (31)$$

Write $f^0(X_m, Y)$ for the polynomial f when interpreted as a polynomial in the two variables X_m, Y with coefficients in the field

$$K = F_q(X_1, \ldots, X_{m-1}). \quad (32)$$

Then $f^0(X_m, Y)$ is irreducible over K. But it may or it may not be absolutely irreducible, i.e., be irreducible over $\overline{K}$. In the next sections we shall assume this as a hypothesis. On the other hand, it will be convenient for later applications to relax (31). Hence we shall make the

HYPOTHESIS H. *The polynomial f is of the type* (30) *with*

$$\deg g_i \leqslant i\psi \qquad (i = 1, \ldots, d), \quad (33)$$

where ψ is a constant $\geqslant 1$. It is separable in Y. Moreover, the polynomial $f^0(X_m, Y)$ is absolutely irreducible.

This hypothesis will be assumed to hold until the end of §11. Later, in §14, we shall show how the general case can be reduced to this special hypothesis.

The following is an immediate consequence of Lemma 11 and Hypothesis H:

LEMMA 13. *Let $\mathfrak{Y}$ be a solution of the equation*

$$f(X_1, \ldots, X_m, \mathfrak{Y}) = 0. \quad (34)$$

Let $a(X_1, \ldots, X_m, Y, Z, W)$ be a non-zero polynomial with coefficients in F_q and with

$$\deg_{X_m} a \leqslant (q/d) - d\psi, \quad (35)$$

$$\deg_Y a \leqslant d - 1, \qquad \deg_W a \leqslant d - 1. \quad (36)$$

Then

$$a(X_1, \ldots, X_m, \mathfrak{Y}, X_m^q, \mathfrak{Y}^q) \neq 0.$$

8. Partial Derivatives

Given a monomial $X_1^{i_1} \cdots X_m^{i_m} Y^u Z^v W^w$, write

$$\Delta = i_m + \psi u, \qquad \nabla = i_1 + \cdots + i_{m-1} + qv.$$

Given a polynomial $a(X_1, \ldots, X_m, Y, Z, W)$, write $\Delta(a)$ for the maximum of Δ for all monomials which have a non-zero coefficient in a, and write $\nabla(a)$ for the maximum of ∇ for all these monomials.

Let $D_1, \ldots, D_m$ be the partial differentiation operators in $F_q(X_1, \ldots, X_m)$. Let $\mathfrak{Y}$ be the algebraic function with (34). Since $\mathfrak{Y}$ is separable over $F_q(X_1, \ldots, X_m)$, each D_i may be extended to a derivation in $F_q(X_1, \ldots, X_m, \mathfrak{Y})$ ([13] §66). In fact we have

$$D_i \mathfrak{Y} = -f_{X_i}(X_1, \ldots, X_m, \mathfrak{Y}) / f_Y(X_1, \ldots, X_m, \mathfrak{Y}) \qquad (i = 1, \ldots, d),$$

where $f_{X_1}, \ldots, f_{X_m}, f_Y$ are the partial derivatives of f. Put

$$\mathfrak{Z} = f_Y(X_1, \ldots, X_m, \mathfrak{Y}).$$

Lemma 14. *Let $a(X_1, \ldots, X_m, Y, Z, W)$ be a polynomial in the $m + 3$ variables $X_1, \ldots, X_m, Y, Z, W$. Suppose $l = l_1 + \cdots + l_m \geqslant 1$. Then*

$$D_1^{l_1} \cdots D_m^{l_m} a(X_1, \ldots, X_m, \mathfrak{Y}, X_m^q, \mathfrak{Y}^q) = a^{(l_1, \ldots, l_m)}(X_1, \ldots, X_m, \mathfrak{Y}, X_m^q, \mathfrak{Y}^q) / \mathfrak{Z}^{2l-1}, \tag{37}$$

where $a^{(l_1, \ldots, l_m)}(X_1, \ldots, X_m, Y, Z, W)$ is a polynomial whose coefficients are linear combinations of the coefficients of a, and which has

$$\begin{gathered} \Delta(a^{(l_1, \ldots, l_m)}) \leqslant \Delta(a) + (2l - 1)\, d\psi, \\ \deg_Z a^{(l_1, \ldots, l_m)} \leqslant \deg_Z a, \qquad \deg_W a^{(l_1, \ldots, l_m)} \leqslant \deg_W a. \end{gathered} \tag{38}$$

Proof. We shall begin with the case $l = 1$. We have

$$\begin{aligned} D_i a(X_1, \ldots, X_m, \mathfrak{Y}, X_m^q, \mathfrak{Y}^q) &= a_{X_i}(X_1, \ldots, X_m, \mathfrak{Y}, X_m^q, \mathfrak{Y}^q) + a_Y(\cdots) D_i \mathfrak{Y} \\ &= b_i(X_1, \ldots, X_m, \mathfrak{Y}, X_m^q, \mathfrak{Y}^q) / \mathfrak{Z}, \end{aligned}$$

where $b_i(X_1, \ldots, X_m, Y, Z, W)$ equals

$$\begin{aligned} &a_{X_i}(X_1, \ldots, X_m, Y, Z, W) f_Y(X_1, \ldots, X_m, Y) \\ &\quad - a_Y(X_1, \ldots, X_m, Y, Z, W) f_{X_i}(X_1, \ldots, X_m, Y). \end{aligned}$$

Thus $\Delta(b_i) \leqslant \Delta(a) + \Delta(f) \leqslant \Delta(a) + d\psi$ $(i = 1,..., m)$, and the degrees of b_i in Z, W do not exceed the corresponding degrees of a. Since clearly the coefficients of b_i are linear combinations of the coefficients of a, the case $l = 1$ of the lemma is true.

To go from l to $l + 1$, we note that

$$\begin{aligned} D_i a^{(l_1,\dots,l_m)}(X_1,\dots, X_m, \mathfrak{Y}, X_m^q, \mathfrak{Y}^q) \\ = b_i^{(l_1,\dots,l_m)}(X_1,\dots, X_m, \mathfrak{Y}, X_m^q, \mathfrak{Y}^q)/\mathfrak{Z} \\ = c_i^{(l_1,\dots,l_m)}(X_1,\dots, X_m, \mathfrak{Y}, X_m^q, \mathfrak{Y}^q)/\mathfrak{Z}^2, \end{aligned}$$

where

$$\Delta(b_i^{(l_1,\dots,l_m)}) \leqslant \Delta(a^{(l_1,\dots,l_m)}) + d\psi,$$

whence

$$\Delta(c_i^{(l_1,\dots,l_m)}) \leqslant \Delta(a^{(l_1,\dots,l_m)}) + 2d\psi.$$

By the case $l = 1$ already proved, we have

$$D_i(1/\mathfrak{Z}^{2l-1}) = (1 - 2l)(D_i\mathfrak{Z})/\mathfrak{Z}^{2l} = d_i(X_1,\dots, X_m, \mathfrak{Y})/\mathfrak{Z}^{2l+1},$$

where $\Delta(d_i) \leqslant \Delta(f_Y) + d\psi \leqslant 2d\psi$. Thus we may put

$$a^{(l_1,\dots,l_i+1,\dots,l_m)} = c_i^{(l_1,\dots,l_m)} + a^{(l_1,\dots,l_m)} d_i,$$

and obtain

$$D_1^{l_1} \cdots D_i^{l_i+1} \cdots D_m^{l_m} a(X_1,\dots, \mathfrak{Y}^q) = a^{(l_1,\dots,l_i+1,\dots,l_m)}(X_1,\dots, \mathfrak{Y}^q)/\mathfrak{Z}^{2l+1}$$

with

$$\Delta(a^{(l_1,\dots,l_i+1,\dots,l_m)}) \leqslant \Delta(a^{(l_1,\dots,l_m)}) + 2d\psi \leqslant \Delta(a) + (2l + 1)\, d\psi.$$

The degrees in Z, W have not increased. Hence the lemma is true for $l + 1$.

Write $a^{(0,\dots,0)} = a$. Then the lemma remains true for $l = 0$ if in (37), (38) we replace $2l - 1$ by $\max(2l - 1, 0)$.

9. Construction of an Algebraic Function

Let $d(X_1,\dots, X_m)$ be the discriminant of the polynomial f when considered as a polynomial in Y with coefficients in $F_q(X_1,\dots, X_m)$. Then by (30) and (33), $d(X_1,\dots, X_m)$ is a polynomial of degree $\leqslant (2d - 2)\psi$. Let σ be the set of m-tuples $(x_1,\dots, x_m)$ with components in F_q having $d(x_1,\dots, x_m) \neq 0$.

Then by Lemma 1 we have

$$q^m - (2d-2)\psi q^{m-1} \leqslant |\sigma| \leqslant q^m. \tag{39}$$

For $(x_1,\ldots,x_m) \in \sigma$, there are precisely d elements $y \in \bar{F}_q$ with $f(x_1,\ldots,x_m,y) = 0$. Let $\tau(x_1,\ldots,x_m)$ consist of all $n = (m+1)$-tuples $(x_1,\ldots,x_m,y)$ with $f(x_1,\ldots,x_m,y) = 0$ and $y \in F_q$, and let $\tau'(x_1,\ldots,x_m)$ consist of the n-tuples $(x_1,\ldots,x_m,y)$ with $f(x_1,\ldots,x_m,y) = 0$ and $y \in \bar{F}_q$, $y \notin F_q$. Then for $(x_1,\ldots,x_m) \in \sigma$,

$$|\tau(x_1,\ldots,x_m)| + |\tau'(x_1,\ldots,x_m)| = d.$$

Let τ and τ', respectively, consist of all n-tuples which belong to $\tau(x_1,\ldots,x_m)$ or to $\tau'(x_1,\ldots,x_m)$ for some $(x_1,\ldots,x_m) \in \sigma$. We clearly have

$$|\tau| + |\tau'| = d|\sigma|. \tag{40}$$

In all that follows, we shall assume that $d \geqslant 2$, and we shall use the abbreviation

$$\mu = ((d-1)/d)^{1/m}. \tag{41}$$

LEMMA. *Suppose*

$$q > m, \qquad q > 2d^2\psi. \tag{42}$$

Suppose M is an integer with

$$M > 2dm, \qquad 20d\psi M^2 < m^2 q. \tag{43}$$

Write

$$N = [\mu(M + 4md)], \tag{44}$$

where $[\cdots]$ denotes the integer part.

Then there is a non-zero polynomial $a(X_1,\ldots,X_m,Y,Z,W)$ with coefficients in F_q and with

$$\deg_{X_m} a < (q/d) - d\psi,$$

$$\nabla(a) \leqslant qN,$$

$$\deg_Y a \leqslant d-1, \qquad \deg_W a \leqslant d-1,$$

such that

$$a^{(l_1,\ldots,l_m)}(x_1,\ldots,x_m,y,x_m^q,y^q) = 0 \quad \text{for} \quad l_1 + \cdots + l_m < M \tag{45}$$

and for every $(x_1,\ldots,x_m,y) \in \tau'$.

Remark. This lemma may be interpreted as follows. The algebraic function $a(X_1,\dots, X_m, \mathfrak{Y}, X_m{}^q, \mathfrak{Y}^q)$ is defined on the surface $f(X_1,\dots, X_m, Y) = 0$, and according to the lemma it vanishes of high order on points $(x_1,\dots, x_m, y) \in \tau'$ of this surface.

Proof. Raising the equation $f(x_1,\dots, x_m, y) = 0$ to the q-th power and observing that $x_1,\dots, x_m$ and the coefficients of f lie in F_q, we get $f(x_1,\dots, x_m, y^q) = 0$. Because of the special form of f we obtain

$$\begin{aligned} 0 &= f(x_1,\dots, x_m, y^q) - f(x_1,\dots, x_m, y) \\ &= (y^q - y)\,((y^{q(d-1)} + y^{q(d-2)}y + \cdots + y^{d-1}) \\ &\quad + g_1(x_1,\dots, x_m)(y^{q(d-2)} + \cdots) + \cdots + g_{d-1}(x_1,\dots, x_m)). \end{aligned}$$

Now if $(x_1,\dots, x_m, y) \in \tau'$, then $y \notin F_q$ and $y^q \neq y$. Hence the second factor above is zero, i.e.,

$$y^{q(d-1)} = -y^{q(d-2)} - \cdots - y^{d-1} - \cdots - g_{d-1}(x_1,\dots, x_m).$$

If we substitute this value for $(y^q)^{d-1}$ in $a^{(l_1,\dots,l_m)}(x_1,\dots, x_m, y, x_m{}^q, y^q)$, we see that (45) becomes

$$b^{(l_1,\dots,l_m)}(x_1,\dots, x_m, y, x_m{}^q, y^q) = 0 \qquad \text{for} \quad l_1 + \cdots + l_m < M, \tag{46}$$

where $b^{(l_1,\dots,l_m)}(X_1,\dots, X_m, Y, Z, W)$ has

$$\deg_W b^{(l_1,\dots,l_m)} \leqslant d - 2,$$

$$\begin{aligned} \Delta(b^{(l_1,\dots,l_m)}) &\leqslant \Delta(a^{(l_1,\dots,l_m)}) + d\psi \leqslant \Delta(a) + (2l+1)\,d\psi \\ &< (q/d) - d\psi + (d-1)\psi + (2l+1)\,d\psi \\ &< (q/d) + (2l+1)\,d\psi, \end{aligned}$$

$$\deg_Z b^{(l_1,\dots,l_m)} \leqslant \deg_Z a \leqslant N.$$

Since $f(x_1,\dots, x_m, y) = 0$, and by (30), (33), one sees by induction on t that

$$y^{d-1+t} = g_1^{(t)}(x_1,\dots, x_m)\,y^{d-1} + \cdots + g_d^{(t)}(x_1,\dots, x_m) \qquad (t = 1, 2,\dots),$$

where

$$\deg g_i^{(t)}(X_1,\dots, X_m) \leqslant (t + i - 1)\psi \leqslant (t + d - 1)\psi \qquad (i = 1,\dots, d).$$

Substituting these values for $y^{d-1+t}(t = 1, 2,...)$ into $b^{(l_1,...,l_m)}(x_1,..., x_m, y, x_m{}^q, y^q)$, we see that (46) is equivalent to

$$c^{(l_1,...,l_m)}(x_1,..., x_m, y, x_m{}^q, y^q) = 0 \qquad \text{for} \quad l_1 + \cdots + l_m < M, \tag{47}$$

where $c^{(l_1,...,l_m)}(X_1,..., X_m, Y, Z, W)$ has

$$\deg_W c^{(l_1,...,l_m)} \leqslant d - 2, \qquad \deg_Y c^{(l_1,...,l_m)} \leqslant d - 1,$$
$$\deg_{X_m} c^{(l_1,...,l_m)} < (q/d) + (2l + 1)\, d\psi,$$
$$\deg_Z c^{(l_1,...,l_m)} \leqslant N.$$

Now finally, since $x_m{}^q = x_m$, the conditions become

$$d^{(l_1,...,l_m)}(x_1,..., x_m, y, y^q) = 0 \qquad \text{for} \quad l_1 + \cdots + l_m < M, \tag{48}$$

where $d^{(l_1,...,l_m)}(X_1,..., X_m, Y, W)$ has

$$\deg_W d^{(l_1,...,l_m)} \leqslant d - 2, \qquad \deg_Y d^{(l_1,...,l_m)} \leqslant d - 1,$$
$$\deg_{X_m} d^{(l_1,...,l_m)} < (q/d) + (2l + 1)\, d\psi + N.$$

The equation in (48) is certainly satisfied if for every $x_1,..., x_{m-1}$ in F_q, the polynomial $d^{(l_1,...,l_m)}(x_1,..., x_{m-1}, X_m, Y, W)$ in X_m, Y, W vanishes identically. The number of coefficients of this polynomial (for fixed $l_1,..., l_m$) is at most

$$(d - 1)\, d((q/d) + (2l + 1)\, d\psi + N + 1)$$
$$< (d - 1)(q + (2l + 2)\, d^2\psi + Nd) = B,$$

say. In order for the above polynomial to vanish identically, the coefficients of $a(X_1,..., X_m, Y, Z, W)$ have to satisfy at most B linear homogeneous equations. In order to have this happen for each $x_1 \in F_q,..., x_{m-1} \in F_q$, we get $\leqslant Bq^{m-1}$ conditions.

The number of m-tuples $l_1,..., l_m$ of non-negative integers with $l_1 + \cdots + l_m < M$ is $\binom{M+m-1}{m}$, hence is $< (M + m)^m/m!$. The sum of $l = l_1 + \cdots + l_m$ over these m-tuples is $< M(M + m)^m/m!$ Taking the sum of $B = B(l_1,..., l_m)$ over all these m-tuples, we get the bound

$$C \leqslant q^{m-1}(d - 1)(M + m)^m\, (m!)^{-1}\, (q + 2d^2\psi + Nd + 2d^2\psi M)$$

for the number of linear conditions on the coefficients of the polynomial a, in order that (48) and hence (45) holds.

Now by (43), (44),

$$2d^2\psi + Nd + 2d^2\psi M < 2d^2\psi + Md + 4md^2 + 2d^2\psi M < 4d^2\psi M,$$

so that

$$C < (d-1)\, q^m (M+m)^m\, (m!)^{-1}\, (1 + (4d^2\psi M/q)). \tag{49}$$

The total number D of coefficients which a polynomial $a(X_1, \ldots, X_m, Y, Z, W)$ of the lemma may have is computed as follows. Since $\nabla(a) \leqslant qN$, every monomial $X_1^{i_1} \cdots X_m^{i_m} Y^u Z^v W^w$ occuring in a has $i_1 + \cdots + i_{m-1} + qv \leqslant qN$. Thus $0 \leqslant v \leqslant N$, and for given v, $i_1 + \cdots + i_{m-1} \leqslant q(N-v)$. For given v, this gives $\binom{q(N-v)+m-1}{m-1}$ possible $(m-1)$-tuples $i_1, \ldots, i_{m-1}$, and the sum over v gives

$$\geqslant \frac{1}{(m-1)!} \sum_{v=1}^{N} (qv)^{m-1} > \frac{q^{m-1}}{m!} (N-1)^m.$$

In view of the further conditions $0 \leqslant i_m < (q/d) - d\psi$, $0 \leqslant u \leqslant d-1$, $0 \leqslant w \leqslant d-1$, the number D of available coefficients satisfies

$$D \geqslant d^2((q/d) - d\psi)\, q^{m-1}(N-1)^m/m!.$$

Now since $d \geqslant 2$, we have $d\mu > 1$, whence $N - 1 \geqslant \mu(M + 4dm) - 2 > \mu(M + 2dm)$, so that

$$D > (d-1)\, q^m (1 - (d^2\psi/q))\, (M + 2dm)^m/m!. \tag{50}$$

The lemma is true if $D > C$, for then the number of available unknowns (i.e., the coefficients of $a(X_1, \ldots, W)$) is greater than the number of homogeneous linear conditions imposed on them. Thus by (49), (50), it will be enough to show that

$$(M + 2dm)^m/(M+m)^m > (1 + (4d^2\psi M/q))\, (1 - (d^2\psi/q))^{-1}. \tag{51}$$

Here the left-hand side is greater than $(1 + (md/(M+m)))^m$, hence is greater than $1 + (m^2 d/2M)$. On the right-hand side, $d^2\psi/q < 1/2$ by (42), so that $(1 - (d^2\psi/q))^{-1} < 1 + (2d^2\psi/q) < 2$. Hence the right-hand side of (51) is less than

$$1 + 8d^2\psi(M/q) + (2d^2\psi/q) < 1 + (10d^2\psi M/q).$$

Thus by (43), the inequality (51) does in fact hold, and the lemma is proved.

10. Construction of a Polynomial

For an m-tuple $(x_1,..., x_m)$ with components in F_q, write

$$\gamma(x_1,..., x_m) = \begin{cases} |\tau'(x_1,..., x_m)| & \text{if } (x_1,..., x_m) \in \sigma, \\ 0 & \text{otherwise.} \end{cases}$$

Lemma 16. *Suppose q, M are as in Lemma 15. There exists a non-zero polynomial $p(X_1,..., X_m)$ with*

$$\deg p \leqslant \mu d M q + (4m + 2\psi)\, d^2 q, \tag{52}$$

such that

$$D_1^{l_1} \cdots D_m^{l_m} p(x_1,..., x_m) = 0 \quad \text{for} \quad l_1 + \cdots + l_m < M\gamma(x_1,..., x_m) \tag{53}$$

for every $(x_1,..., x_m)$.

Proof. Put

$$p(X_1,..., X_m) = \mathfrak{N}(a(X_1,..., X_m, \mathfrak{Y}, X_m^q, \mathfrak{Y}^q)), \tag{54}$$

where $\mathfrak{N}$ denotes the norm from $F_q(X_1,..., X_m, \mathfrak{Y})$ to $F_q(X_1,..., X_m)$, and where $a(X_1,..., X_m, Y, Z, W)$ is the polynomial of Lemma 15. Let $\mathfrak{Y} = \mathfrak{Y}^{(1)},..., \mathfrak{Y}^{(d)}$ be the conjugates of $\mathfrak{Y}$. The right-hand side of (54) is symmetric in $\mathfrak{Y}^{(1)},..., \mathfrak{Y}^{(d)}$ and of degree $\leqslant (d-1)(1+q)$ in each $\mathfrak{Y}^{(i)}$. Such a symmetric function is by Lemma 10 a polynomial of degree $\leqslant (q+1)(d-1)$ in the elementary symmetric functions of $\mathfrak{Y}^{(1)},..., \mathfrak{Y}^{(d)}$, hence is by (33) a polynomial in $X_1,..., X_m$ of degree $\leqslant (q+1)(d-1)\, d\psi$. Since $\deg_{X_m} a < q/d$ and since $\nabla(a) \leqslant Nq$, it follows that

$$\begin{aligned} \deg p &\leqslant (q+1)(d-1)\, d\psi + q + Nqd \\ &< qd^2\psi + q + \mu(M + 4md)\, qd \\ &< \mu d\, Mq + (4m + 2\psi)\, d^2 q \end{aligned}$$

by (42), (44). Thus (52) holds. The polynomial p is non-zero by Lemma 13.

Each derivation D_i is uniquely extended to $F_q(X_1,..., X_m, \mathfrak{Y}^{(1)},..., \mathfrak{Y}^{(d)})$. Since $p(X_1,..., X_m)$ is the product of $a(X_1,..., X_m, \mathfrak{Y}^{(i)}, X_m^q, \mathfrak{Y}^{(i)q})$ $(i = 1,..., d)$, we have by Lemma 14,

$$\begin{aligned} &D_1^{l_1} \cdots D_m^{l_m} p(X_1,..., X_m) \\ &\quad = \sum c(l_{ij}) \prod_{i=1}^{d} (a^{(l_{i1},..., l_{im})}(X_1,..., X_m, \mathfrak{Y}^{(i)}, X_m^q, \mathfrak{Y}^{(i)q})/\mathfrak{Z}_i^{\lambda_i}), \end{aligned} \tag{55}$$

where the sum is over non-negative l_{ij} $(1 \leqslant i \leqslant d,\ 1 \leqslant j \leqslant m)$ with

$$l_{1j} + \cdots + l_{dj} = l_j \qquad (j = 1,\ldots, m), \tag{56}$$

where the $c(l_{ij})$ are constants depending only on $l_{11},\ldots, l_{dm}$, where $\mathfrak{Z}_i = f_Y(X_1,\ldots, X_m, \mathfrak{Y}^{(i)})$ and where $\lambda_i = \max(0, 2(l_{i1} + \cdots + l_{im}) - 1)$ $(i = 1,\ldots, d)$.

Now suppose that $(x_1,\ldots, x_m) \in \sigma$, and that $y_1,\ldots, y_d$ are the roots of $f(x_1,\ldots, x_m, y) = 0$. Since $z_i = f_Y(x_1,\ldots, x_m, y_i) \neq 0$ $(i = 1,\ldots, d)$, (55) yields

$$D_1^{l_1} \cdots D_m^{l_m} p(x_1,\ldots, x_m)$$

$$= \sum c(l_{ij}) \prod_{i=1}^{d} (a^{(l_{i1},\ldots,l_{im})}(x_1,\ldots, x_m, y_i, x_m^q, y_i^q)/z_i^{\lambda_i}). \tag{57}$$

Suppose $y_1,\ldots, y_\gamma$ are $\notin F_q$ and $y_{\gamma+1},\ldots, y_d$ are $\in F_q$, so that

$$\gamma = |\tau'(x_1,\ldots, x_n)| = \gamma(x_1,\ldots, x_n).$$

Now let $l_1 + \cdots + l_m < M\gamma(x_1,\ldots, x_m) = M\gamma$. We have $l_{1j} + \cdots + l_{\gamma j} \leqslant l_j$ $(j = 1,\ldots, m)$ by (56), so that there is an i, $1 \leqslant i \leqslant \gamma$, with

$$l_{i1} + \cdots + l_{im} \leqslant (l_1 + \cdots + l_m)/\gamma < M.$$

Since $(x_1,\ldots, x_m, y_i) \in \tau'$, Lemma 15 gives

$$a^{(l_{i1},\ldots,l_{im})}(x_1,\ldots, x_m, y_i, x_m^q, y_i^q) = 0.$$

Hence every summand on the right-hand side of (57) is zero, and the lemma is established.

11. A Lower Bound for the Number of Zeros, Derived under Hypothesis H

Lemma 17. *Suppose $k(X_1,\ldots, X_m)$ is a non-zero polynomial with coefficients in F_q and of degree $\leqslant e$. Suppose B is a constant $\leqslant p$, and with every $(x_1,\ldots, x_m)$ with components in F_q there is associated an integer $\beta(x_1,\ldots, x_m)$ with*

$$0 \leqslant \beta(x_1,\ldots, x_m) \leqslant B. \tag{58}$$

Suppose that

$$D_1^{l_1} \cdots D_m^{l_m} k(x_1, \ldots, x_m) = 0 \qquad \text{for} \quad l_1 + \cdots + l_m < \beta(x_1, \ldots, x_m)$$

for every $(x_1, \ldots, x_m)$. *Then*

$$\sum_{(x_1, \ldots, x_m) \in F_q^m} \beta(x_1, \ldots, x_m) \leqslant e_m ,$$

where

$$e_m = e(q^{m-1} + B(q^{m-2} + \cdots + 1)).^5$$

Proof. For $m = 1$, we have $D^l k(x) = 0$ for $l = 0, 1, \ldots, \beta(x) - 1$, and therefore $k(X)$ is divisible by $(X - x)^{\beta(x)}$. (This is only true since $\beta \leqslant B \leqslant p$, where p is the characteristic; it would not be true without this condition!) Thus $k(X)$ has degree at least $\sum \beta(x)$, and it follows that $\sum \beta(x) \leqslant e = e_1$.

The induction from $m - 1$ to m is as follows. We shall denote hyperplanes (not necessarily through the origin) of F_q^m by the letter H. With each hyperplane H we associate a linear form $f_H(X_1, \ldots, X_m)$, such that H consists of the zeros of f_H. Put

$$\beta(H) = \sum_{(x_1, \ldots, x_m) \in H} \beta(x_1, \ldots, x_m).$$

Now if we parametrize H and substitute the expressions for $X_1, \ldots, X_m$ into $k(X_1, \ldots, X_m)$, we obtain a polynomial in $m - 1$ variables of degree $\leqslant e$. Thus by our inductive hypothesis we see that if $\beta(H) > e_{m-1}$, then k vanishes identically on H, whence is divisible by f_H. Hence there are at most e hyperplanes H with $\beta(H) > e_{m-1}$. Every hyperplane H has $\beta(H) \leqslant Bq^{m-1}$. Therefore

$$\sum_H \beta(H) \leqslant \sum_H e_{m-1} + \sum_{\substack{H \\ \beta(H) > e_{m-1}}} Bq^{m-1}$$

$$\leqslant e_{m-1} q(q^{m-1} + \cdots + q + 1) + eBq^{m-1},$$

since the number of hyperplanes is $q(q^{m-1} + \cdots + q + 1)$. The number of hyperplanes through a given point is $q^{m-1} + \cdots + q + 1$. Hence

$$\sum_{(x_1, \ldots, x_m)} \beta(x_1, \ldots, x_m) = (q^{m-1} + \cdots + 1)^{-1} \sum_H \beta(H) < e_{m-1} q + eB = e_m .$$

[5] The inner sum is zero if $m = 1$.

LEMMA 18. *Suppose Hypothesis H holds, and suppose*

$$q > 20md^5\psi(16m + 8\psi)^2. \tag{59}$$

Then the number A of solutions of (1) *satisfies*

$$A \geqslant q^m/(5m). \tag{60}$$

Proof. First consider the case when F_q is a prime field, i.e., when $q = p$. Put

$$M = (16m + 8\psi)\, d^2m. \tag{61}$$

Then q, M satisfy the conditions of Lemma 15, 16. We are going to apply Lemma 17 to the polynomial $p(X_1, \ldots, X_m)$ of Lemma 16, with $\beta(x_1, \ldots, x_m) = M\gamma(x_1, \ldots, x_m)$ and with $B = Md$. Then $B \leqslant p$ in view of (59), (61) and $q = p$. We obtain

$$\begin{aligned} e_m &= e(q^{m-1} + B(q^{m-2} + \cdots + 1)) \\ &\leqslant (\mu dMq + (4m + 2\psi)\, d^2q)(q^{m-1} + 2Mdq^{m-2}) \\ &= \mu dMq^m(1 + (4m + 2\psi)\, d(\mu M)^{-1})(1 + (2Md/q)). \end{aligned}$$

Now $1/2 \leqslant 1 - (1/d) \leqslant \mu = (1 - (1/d))^{1/m} \leqslant 1 - (1/dm)$. We have $(4m + 2\psi)\, d(\mu M^{-1}) \leqslant 1/(2dm)$ by (61), and $2Md/q \leqslant 1/(20\, dm)$ by (59), (61). Since $(1 - (1/dm))\,(1 + (1/2\, dm))\,(1 + (1/20\, dm)) < 1 - (2/5\, dm)$, we obtain

$$e_m < (d - (2/5m))\, Mq^m.$$

Lemma 17 yields

$$\begin{aligned} \sum_{(x_1,\ldots,x_m)} \gamma(x_1, \ldots, x_m) &= M^{-1} \sum_{(x_1,\ldots,x_m)} \beta(x_1, \ldots, x_m) \\ &\leqslant M^{-1}e_m < (d - (2/5m))\, q^m. \end{aligned} \tag{62}$$

Now A, the number of solutions of (1), satisfies

$$A \geqslant |\tau| = d\,|\sigma| - |\tau'| > d(q^m - 2d\psi\, q^{m-1}) - \sum \gamma(x_1, \ldots, x_m),$$

by virtue of (39), (40). We obtain

$$A > (2/5m)\, q^m - 2d^2\psi q^{m-1} > q^m/(5m)$$

by (59) and (62).

All this was proved under the assumption that $q = p$. In the general case one has to replace the operators D_j^l by new operators $E_j^{(l)}$ where

$$E_j^{(l)}(X_1^{i_1} \cdots X_j^{i_j} \cdots X_m^{i_m}) = \binom{l}{i_j} X_1^{i_1} \cdots X_j^{i_j - l} \cdots X_m^{i_m}.$$

Lemma 17 remains true without the condition $B \leqslant p$ if $D_1^{l_1} \cdots D_m^{l_m}$ is replaced by $E_1^{(l_1)} \cdots E_m^{(l_m)}$. There are some technical details to be overcome in order to define these new operators on $F_q(X_1, \ldots, X_m, \mathfrak{Y})$, etc. A detailed exposition of these operators (in the case $m = 1$) is given in [7], §8.

COROLLARY. *Suppose Hypothesis H holds and q satisfies* (59) *and*

$$q > 25nd^2q^{1/2} + 5n\alpha(q, d). \tag{63}$$

Then

$$A > q^{n-1} - (\alpha(q, d) + 5d^2)\, q^{n-2}.$$

Proof. We may assume $d \geqslant 2$. Lemma 7 is applicable by (63). If it were true with $\nu_0 = 0$, then $q^{n-1}/(5n) < A < (\alpha(q, d) + 5d^2)\, q^{n-2}$, contradicting (63). Hence $\nu_0 \geqslant 1$, and the desired conclusion follows.

12. COUNTING SUBGROUPS AND SUBFIELDS

The results of this section are not new.

Write $\log_2$ for the logarithm to the base 2, and $[\cdots]$ for the integer part.

LEMMA 19. *Let G be a group and H a subgroup of index h. Then the number of subgroups J with*

$$H \subseteq J \subseteq G \tag{64}$$

is

$$\leqslant h^{[\log_2 h]} = \lambda(h), \quad \textit{say}.$$

Proof. Let $C_1, \ldots, C_h$ be the right cosets of H in G. Every group J as above is a union of such cosets. Write $J(C_{i_1}, \ldots, C_{i_t})$ for the subgroup of G generated by H, $C_{i_1}, \ldots, C_{i_t}$. Every J may be written as

$$J = J(C_{i_1}, \ldots, C_{i_t})$$

for some t and some cosets $C_{i_1}, \ldots, C_{i_t}$. If t is chosen minimal, then in $H \subseteq J(C_{i_1}) \subseteq \cdots \subseteq J(C_{i_1}, \ldots, C_{i_t})$ each extension is of degree $\geqslant 2$, whence the index of H in $J(C_{i_1}, \ldots, C_{i_t})$ at least 2^t. Therefore $t \leqslant t_0 = [\log_2 h]$.

In fact we may always take $t = t_0$. The number of possibilities for each i_j ($j = 1,..., t_0$) is h, so that we obtain the upper bound h^{t_0}.

LEMMA 20. *Let K be a field and L a separable algebraic extension of degree h. Then the number of fields M with*

$$K \subseteq M \subseteq L \tag{65}$$

is $\leqslant \lambda(h)$.

Proof. This follows from Galois Theory: Let N be the smallest normal extension of K containing L. Let G be the Galois group of N over K, and H the subgroup consisting of elements which leave L fixed. Then H is a subgroup of index h. There is a 1–1 correspondence between fields M with (65) and subgroups J of G with (64). Our assertion now follows from Lemma 19.

We shall assume again that $d \geqslant 2$. Put $u = [4 \log d]$, let $P(1) = 2,..., P(u)$ be the first u primes, and set

$$\psi = P(u). \tag{66}$$

Let Ω be the set of polynomials over F_q of the type

$$r(X) = \sum a_P X^P, \tag{67}$$

where the sum is over primes $P \leqslant \psi$ which are not divisors of d.

LEMMA 21. *Suppose $q > d$. Then $| \Omega | > \lambda(d)$.*

Proof. Since the number of prime factors of d is $\leqslant [\log_2 d]$, the number of primes $P \leqslant \psi = P(u)$ which do not divide d is $\geqslant u - [\log_2 d] \geqslant 2[\log_2 d] - [\log_2 d]$. For each such prime P, the number of possibilities for a_P is q, so that

$$| \Omega | \geqslant q^{[\log_2 d]} > d^{[\log_2 d]} = \lambda(d).$$

13. On Absolutely Irreducible Polynomials

Given a field K and a field extension L, the *algebraic closure of K in L* consists of the elements of L which are algebraic over K. The algebraic closure is a field K^* with $K \subseteq K^* \subseteq L$.

LEMMA 22. *Suppose $f(X_1,..., X_m, Y)$ is a polynomial with coefficients in K, which is irreducible over K and separable and of degree $d > 0$ in Y.*

Let $\mathfrak{Y}$ be a quantity with $f(X_1, \ldots, X_m, \mathfrak{Y}) = 0$. Let L be the field $K(X_1, \ldots, X_m, \mathfrak{Y})$. Then the algebraic closure K^* of K in L is a separable algebraic extension of K, with $[K^*: K]$ a divisor of d. The polynomial f is absolutely irreducible if and only if $K^* = K$.

Proof. In the chain

$$K(X_1, \ldots, X_m) \subseteq K^*(X_1, \ldots, X_m) \subseteq L = K^*(X_1, \ldots, X_m, \mathfrak{Y}), \quad (68)$$

the field on the right is a separable algebraic extension of degree d of the field on the left. Hence $K^*(X_1, \ldots, X_m)$ is a separable algebraic extension of $K(X_1, \ldots, X_m)$ of a degree dividing d, and K^* is a separable algebraic extension of K of a degree dividing d.

If f is absolutely irreducible, then it is irreducible over K^*, so that $[K^*(X_1, \ldots, X_m, \mathfrak{Y}): K^*(X_1, \ldots, X_m)] = d$, and therefore $K^*(X_1, \ldots, X_m) = K(X_1, \ldots, X_m)$ and $K^* = K$.

In general, let f_1 be an irreducible factor of f over $\overline{K}$. Assume that in some lexicographic ordering of the coefficients, the first non-zero coefficient of f_1 is 1. Then the same is true for every power of f_1. Let K_1 be obtained from K by adjoining the coefficients of f_1. Let a be the smallest integer such that the coefficients of f_1^a are separable over K. Then f_1^b has separable coefficients precisely if a divides b. Now $f_1^{p^e}$ does have separable coefficients for some power p^e of the characteristic p. Hence a is a divisor of such a power, hence is itself a power of p. We have

$$K \subseteq K_1^s \subseteq K_1,$$

where K_1^s is obtained from K be adjoining the coefficients of $g = f_1^a$, and where K_1 is a purely inseparable extension of K_1^s. The polynomial $g = f_1^a$ has coefficients in K_1^s and is irreducible over K_1^s, since its proper divisors don't have separable coefficients. Now $g = f_1^a$ divides f^a, hence g divides f. Write $k = [K_1^s: K]$ and let $f_1, f_2, \ldots, f_k$ be the (distinct) conjugates of f_1 over K. Then each power f_i^a $(i = 1, \ldots, k)$ divides f, whence $h = (f_1 f_2 \cdots f_k)^a$ divides f. Since h has coefficients in K and since f is irreducible over K, $f = ch$ with a constant c. If a is a positive power of p, then f is a polynomial in $X_1^p, \ldots, X_m^p, Y^p$, which contradicts the separability of f in Y. Hence $a = 1$ and $f = cf_1 f_2 \cdots f_k$. Each polynomial f_i is of degree d/k in Y, so that $\mathfrak{Y}$ is of degree d/k over $K_1(X_1, \ldots, X_m)$. Thus in

$$K(X_1, \ldots, X_m) \subseteq K_1(X_1, \ldots, X_m) \subseteq K_1(X_1, \ldots, X_m, \mathfrak{Y}),$$

the first extension is of degree k and the second is of degree d/k. Hence the

total extension is of degree d, and $K_1(X_1,\dots, X_m, \mathfrak{Y}) = K(X_1,\dots, X_m, \mathfrak{Y}) = L$ and $K_1 \subseteq L$. Since K_1 is algebraic over K, $K_1 \subseteq K^*$.

Now if $K = K^*$, then $K_1 = K$, whence $k = 1$, whence $f = cf_1$. Hence f is absolutely irreducible.

14. Conclusion of the Proof

Let $f(X_1,\dots, X_m, Y)$ be the polynomial of the Theorem. We may assume that it is of the type (30), (31), and that it is separable in Y. Let $\mathfrak{Y}$ satisfy (34), and let

$$L = F_q(X_1,\dots, X_m, \mathfrak{Y}).$$

Write $K_0 = F_q$. Then by Lemma 22, and since f is absolutely irreducible, $K_0^* = K_0$, where K_0^* is the algebraic closure of K_0 in L. Put

$$K_i = F_q(X_1,\dots, X_i) \qquad (i = 1,\dots, m).$$

Hypothesis H is that the polynomial $f^0(X_m, Y)$ (i.e., the polynomial f as a polynomial in X_m, Y with coefficients in K_{m-1}) is absolutely irreducible. By Lemma 22 this is true precisely if K_{m-1} is algebraically closed in $K_{m-1}(X_m, \mathfrak{Y}) = L$. Thus Hypothesis H means that

$$K_{m-1}^* = K_{m-1}. \tag{69}$$

But of course in general this need not be true. We know that $K_0^* = K_0$, but we don't even know whether $K_1^* = K_1$.

For every $r \in \Omega$, put

$$X_1^{(r)} = X_1 + r(X_m).$$

Lemma 23. *There is an $r \in \Omega$ with $(K_0(X_1^{(r)}))^* = K_0(X_1^{(r)})$.*

Proof. The fields $(K_0(X_1^{(r)}))^*(X_1,\dots, X_m)$ lie between K_m and L. By Lemma 20, there are at most $\lambda(d)$ fields between K_m and L. Hence by Lemma 21, there are two distinct polynomials $r, s \in \Omega$ with

$$(K_0(X_1^{(r)}))^*(X_1,\dots, X_m) = (K_0(X_1^{(s)}))^*(X_1,\dots, X_m).$$

Since $X_2,\dots, X_{m-1}$ are algebraically independent over $K_0(X_1, X_m)$, we obtain

$$(K_0(X_1^{(r)}))^*(X_1, X_m) = (K_0(X_1^{(s)}))^*(X_1, X_m). \tag{70}$$

Now $\mathfrak{Y}$ is separable algebraic of degree d over $K_0(X_1^{(r)})(X_2,\ldots,X_m)$. Applying Lemma 22 to the ground field $K_0(X_1^{(r)})$, we see that $(K_0(X_1^{(r)}))^*$ is separable algebraic over $K_0(X_1^{(r)})$ of some degree d_r dividing d. Say $(K_0(X_1^{(r)}))^* = K_0(X_1^{(r)}, \mathfrak{Z}_r)$, and let $g_r(X_1^{(r)}, Z)$ be the defining polynomial of $\mathfrak{Z}_r$ over $K_0(X_1^{(r)})$. In view of $K_0^* = K_0$, K_0 is algebraically closed in $K_0(X_1^{(r)}, \mathfrak{Z}_r)$, and hence $g_r(X_1^{(r)}, Z)$ is absolutely irreducible by Lemma 22. Similarly $(K_0(X_1^{(s)}))^* = K_0(X_1^{(s)}, \mathfrak{Z}_s)$, where $\mathfrak{Z}_s$ is separable algebraic over $K_0(X_1^{(s)})$ of some degree d_s dividing d and where the defining polynomial for $\mathfrak{Z}_s$ is absolutely irreducible. Now by the construction of Ω, the degree of $r(Y) - s(Y)$ is a prime not dividing d, hence is a number coprime to $d_r d_s$. Thus Lemma 9 is applicable and

$$[K_0(X_1, X_m, \mathfrak{Z}_r, \mathfrak{Z}_s): K_0(X_1, X_m)] = d_r d_s,$$

while by (70), $K_0(X_1, X_m, \mathfrak{Z}_r) = K_0(X_1, X_m, \mathfrak{Z}_s)$ is of degree $d_r = d_s$ over $K_0(X_1, X_m)$. Hence $d_r = d_s = d_r d_s$, whence $d_r = d_s = 1$, whence $(K_0(X_1^{(r)}))^* = K_0(X_1^{(r)})$.

Now let the polynomial r of the lemma be fixed and put

$$\hat{X}_1 = X_1^{(r)} = X_1 + r(X_m). \tag{71}$$

Introduce the polynomial

$$\hat{f}_1(\hat{X}_1, X_2,\ldots, X_m, Y) = f(\hat{X}_1 - r(X_m), X_2,\ldots, X_m, Y).$$

Then $\hat{f}_1$ is again of degree d in Y and is again separable in Y. We have $\hat{f}_1(\hat{X}_1, X_2,\ldots, X_m, \mathfrak{Y}) = 0$ and $L = F_q(\hat{X}_1, X_2,\ldots, X_m, \mathfrak{Y})$. The polynomial $\hat{f}_1$ is absolutely irreducible and its number of zeros is equal to the number of zeros of f. Putting $\hat{K}_1 = K_0(\hat{X}_1)$, we have $\hat{K}_1^* = \hat{K}_1$. Hence we see that *after a suitable substitution* (71), *we have*

$$\hat{K}_1^* = \hat{K}_1.$$

We now interpret $\hat{f}_1$ as a polynomial in $X_2,\ldots, X_m, Y$ with coefficients in $\hat{K}_1$. If $m \geqslant 3$, we may repeat the process and we see that after a substitution $\hat{K}_2 = X_2 + r_2(X_m)$, the field $\hat{K}_2 = K_0(\hat{X}_1, \hat{X}_2)$ will have $\hat{K}_2^* = \hat{K}_2$. After $m - 1$ substitutions we finally get $\hat{K}_{m-1}^* = \hat{K}_{m-1}$.

But after $m - 1$ substitutions the polynomial obtained is of the type

$$\hat{f}_{m-1} = f(X_1 - r_1(X_m),\ldots, X_{m-1} - r_{m-1}(X_m), X_m, Y),$$

where f was the original polynomial. Since f was of the type (30) with (31) and since $\deg r_i(X_m) \leqslant \psi$ $(i = 1,\ldots, m-1)$, the new polynomial will be of the type (30), (33) with $\psi = \psi(d) = P([4 \log d])$ according to (66). Hence Hypothesis H will hold with this value of ψ.

Now suppose (2) holds, and set $\alpha(q, d) = \alpha_1 = (d-1)(d-2)q^{1/2} + d$. Then q satisfies (59) and (63), and we may apply the Corollary to Lemma 18. Thus

$$A > q^{n-1} - (\alpha(q, d) + 5d^2)\, q^{n-2}$$
$$\geqslant q^{n-1} - (d-1)(d-2)\, q^{n-(3/2)} - 6d^2 q^{n-2},$$

i.e., (3).

Now on the other hand, suppose that (2), (2′) hold, and apply our estimates with $\alpha(q, d) = \alpha_2 = 2d^3 q^{1/2}$. Then again (59) and (63) are true, and the Corollary to Lemma 18 yields

$$A > q^{n-1} - (\alpha(q, d) + 5d^2)\, q^{n-2} \geqslant q^{n-1} - 3d^3 q^{n-(3/2)},$$

i.e., (3′).

References

1. E. Bombieri, "Counting Points on curves over finite Fields [d'après S. A. Stepanov]." Seminaire Bourbaki, 25e année, 1972/73, n° 430.
2. Z. I. Borevich and I. R. Shafarevich, "Number Theory." Academic Press, New York and London, 1966 (translated from the Russian 1964 edit.).
3. B. Dwork. On the rationality of the zeta function of an algebraic variety, *Am. J. Math.* **82** (1960), 631–648.
4. B. Dwork, On the zeta function of a hypersurface. *Inst. Hautes Études Sci. Publ. Math.* **12** (1962), 5–68.
5. S. Lang and A. Weil, The number of points of varieties in finite fields. *Amer. J. Math.* **76** (1954), 819–827.
6. L. B. Nisnevich, On the number of points of an algebraic variety in a finite prime field, *Dokl. Akad. Nauk SSSR* **99** (1954), 17–20 (in Russian).
7. Wolfgang M. Schmidt, Zur Methode von Stepanov, *Acta Arith.* **24** (1973), 247–267.
8. S. A. Stepanov. The number of points of a hyperelliptic curve over a prime field, *Izv. Akad. Nauk SSSR Ser. Mat.* **33** (1969), 1171–1181 (in Russian).
9. S. A. Stepanov, Elementary method in the theory of congruences for a prime modulus, *Acta Arith.* **17** (1970), 231–247.
10. S. A. Stepanov, Estimates of rational trigonometric sums with prime denominators, *Trudy Akad. Nauk* **62** (1971), 346–371 (in Russian).
11. S. A. Stepanov, An elementary proof of the Hasse–Weil Theorem for hyperelliptic curves, *J. Number Theory* **4** (1972), 118–143.
12. S. A. Stepanov, Congruences in two variables, *Izv. Akad. Nauk SSSR Ser. Mat.* **36** (1972), 683–711.
13. B. L. van der Waerden, "Algebra I," 4th edit. Berlin–Göttingen–Heidelberg, 1955.
14. A. Weil, Sur les courbes algébriques et les variétés qui s'en déduisent, *Act. Sci. Ind.* Paris (1948), 1041.